

Transforming in the Digital Tide: An In-depth Analysis of the Impact of E-Commerce on Brick-and-Mortar Retail in China, 2019-2023

Qingyue Chen 22-737-639
Shiyang Jin 22-737-712
Yuru Jin 21-749-650
Zhengyu Duan 22-736-904

December 5, 2023

Contents

1	Introduction	2
2	Literature Review	2
3	Assumption	3
4	Prediction	3
5	Data Section	3
6	Descriptive analysis	5
7	Statistic analysis	6
7.1	Seasonal analysis	6
7.2	t-test	7
7.3	Regression analysis	8
7.4	Deseasonalized regression	9
8	Conclusion	9

1 Introduction

The rapid expansion and integration of e-commerce in China has significantly changed the retail landscape in recent years. This report presents an in-depth analytical exploration of this transformation, focusing on how the burgeoning e-commerce sector influences and reshapes traditional brick-and-mortar retail. The comprehension of global retail trends necessitates a thorough awareness of the dynamics in China, which currently has a prominent position in Internet commerce. Our analysis begins with a comprehensive overview of the retail market in China, laying the groundwork to appreciate the scale and pace of change. We delve into the historical data on retail sales, examining online and offline channels to identify trends, patterns, and significant shifts in consumer behaviour. The aforementioned historical framework serves as a foundation upon which the contemporary emergence of e-commerce can be situated. We then turn our attention to the core of this report: the interplay between e-commerce and traditional retail. Here, we dissect the growth rates of online sales against their offline counterparts, revealing the impact of digital platforms on market share and consumer preferences and how traditional retailers adapt to this new era. This analysis seeks to illuminate the nuances of the transition from physical stores to digital platforms, highlighting areas of synergy, competition, and transformation. While our investigation touches on topics pertinent to various stakeholders in the retail industry, it is essential to clarify that this report's primary objective is analytical. We aim to offer a detailed, data-driven examination of the evolving retail landscape in China. Our focus is on unpacking the complex dynamics at play and providing a narrative that is rich in detail and insights.

In conclusion, this report seeks to contribute to the broader dialogue on the future of retail in the digital age. Through a comprehensive analysis of the Chinese market, recognised for its pioneering advancements in e-commerce and digital consumer interaction, our objective is to elucidate discernible patterns and trends that can shape and impact the worldwide retail industry.

2 Literature Review

The e-commerce landscape has been transformative since Amazon's inception in 1994. Beginning as a bookseller, Amazon expanded into a diverse global marketplace, achieving sales of \$136 billion by 2016. In parallel, the founding of Alibaba.com in China in 1999 marked the beginning of a significant shift in the Asian e-commerce market. Alibaba's 2018 Singles' Day sale, amassing \$30.8 billion in transactions, underscored the global impact of e-commerce, dwarfing the sales of Black Friday and Cyber Monday in the United States[LWZ19].

This evolution of e-commerce as a novel business model has profoundly affected consumer behaviours and consumption patterns[FTZZ18]. In developed countries, the discourse around e-commerce has focused on its welfare implications—ranging from reduced consumer search costs and increased product variety to its effects on traditional retail sectors, such as the decline seen in physical bookstores.

The growth trajectory of e-commerce in China is particularly noteworthy in this global narrative. By 2017, China boasted 772 million internet users, with 533 million participating in online shopping. The annual e-commerce trade volume soared from RMB 930 billion in 2004 to RMB 29,160 billion in 2017, highlighting a compound annual growth rate of 30%. This explosive growth was mirrored in online retail sales, which increased from RMB 125.7 billion in 2008 to RMB 5,155.6 billion in 2016. Internationally, China's contribution to global e-commerce transactions rocketed from less than 1% a decade ago to over 40%, surpassing the combined totals of economic powerhouses like France, Germany, Japan, the UK, and the US[LWZ19].

Yet, within China, the development of e-commerce reveals significant regional and urban-rural disparities. Central provinces such as Beijing, Shanghai, Zhejiang, and Guangdong recorded high percentages of online retail sales in 2015, while in contrast, less than 2% of retail sales were online in nine inland provinces. This dichotomy extends to the urban-rural divide, with 72.6% of the national total urban internet users compared to 27.4% in rural areas[Min16]. However, rural online retail transactions have grown more rapidly, increasing from RMB 353 billion in 2015 to RMB 895 billion in 2016[Min16].

This backdrop sets the stage for our investigation into how the e-commerce surge impacts traditional brick-and-mortar retail dynamics in China. As we delve into the nuances of this relationship, we aim to unravel the complexities of a market in transition, where the traditional and the digital coexist and compete, shaping the future of retail in one of the world's most vibrant economies.

3 Assumption

- **Data Reliability:** Assuming the data for e-commerce and traditional retail in China is accurate and representative of market trends.
- **Consistent Consumer Behavior:** Assuming consumer preferences remained stable, focusing solely on the impact of e-commerce on sales dynamics.
- **Broad Representation:** Assuming the data encompasses diverse geographic and demographic segments across China.
- **Technology Access:** Assuming widespread access to and usage of e-commerce platforms across different consumer segments.
- **Clear Retail Segmentation:** Assuming a distinct separation between online and offline retail channels for analysis purposes.

4 Prediction

- The increasing negative correlation between online and offline retail channels in China signals a consumer shift towards digital platforms, impacting traditional brick-and-mortar store traffic.
- The COVID-19 pandemic has accelerated the move to e-commerce, likely causing enduring changes in consumer behavior and underscoring the need for retail agility and digital readiness.

5 Data Section

Data Sources and Collection

Sources

Our primary data source is the official website of the National Bureau of Statistics of China, which provides monthly growth figures for online and total consumption. This rich dataset offers an in-depth view of the evolving retail landscape in China from 2019 to 2023, encompassing various product categories and retail formats.

Data Collection

The collection process involved the use of web scraping techniques to efficiently gather large volumes of data. By deploying Python-based scraping tools, we were able to automate the extraction of relevant data, ensuring accuracy and timeliness.

Data Description

Overview

The dataset spans from 2019 to 2023, detailing monthly consumption amounts for both online and offline channels. It includes relative growth rates, breaking down offline sales into different categories to reveal trends and shifts in consumer behavior.

Key Insights

- The data reveals a dynamic interplay between online and offline retail channels, with noticeable trends and shifts over the examined period.
- Offline sales are categorized into various product types, offering insights into consumer preferences and market dynamics.

Data Processing

Approach

In processing the data, our aim was to create a cohesive and comprehensive dataset that accurately reflects market dynamics. This involved:

1. **Data Integration and Summarization:** Combining various datasets to form a unified view of the retail landscape.
2. **Handling Data Shortages:** For January of each year, where data is typically scarce, we applied an averaging technique. The average of February's data was used to estimate January's figures, with February data remaining unchanged in cumulative metrics.
3. **Correlation Analysis:** Investigating the negative correlation between online and offline sales, particularly under the influence of COVID-19, and its impact on the overall retail sales.

Data Crawling with Python

Methodology

Our data collection from the National Bureau of Statistics of China was executed through a Python-based web scraping approach. Key steps of the methodology included:

1. **Data Extraction:** Crafting HTTP POST requests with appropriate headers and parameters to access the targeted datasets. The data, spanning from 2019 to 2023, included detailed retail sales figures and growth rates.
2. **Parsing and Cleaning:** Employing custom functions to parse JSON responses, extract relevant data points, and handle missing or incomplete data for consistency and accuracy.
3. **Data Structuring:** Converting the scraped data into a structured format using pandas DataFrames, facilitating easy manipulation and analysis in subsequent stages.

This streamlined and ethical web scraping process ensured the collection of comprehensive and reliable data, forming the backbone of our analysis.

Key Considerations in Data Handling

1. **Integration of Diverse Data Sources:** The challenge was to merge data from different formats and sources into a single, coherent dataset.
2. **Handling Seasonal Shortages:** The approach to address data scarcity in January each year was crucial to maintain the integrity of the analysis.
3. **Cumulative Data Considerations:** In dealing with cumulative data, special attention was paid to ensure that modifications (like the averaging for January) did not skew the overall trends.
4. **Online-Offline Correlation and COVID-19 Impact:** It was essential to understand and articulate how the pandemic influenced retail dynamics, affecting both online and offline channels.

Conclusion

This data-centric approach lays the foundation for a comprehensive analysis of the evolving retail landscape in China, highlighting key trends and patterns that have implications for the global retail sector.

6 Descriptive analysis

First, we conducted a descriptive statistical analysis of the obtained data. Use python to load the data set and then clean the data. Since we only have the monthly cumulative value of online sales, we need to calculate the monthly online sales through the cumulative value. Through the cumulative value of total retail sales and online retail sales, we can also calculate the value of monthly offline sales.

This graph shows the monthly online retail sales in goods over a period from 2019 to 2023. The sales values, measured in millions (M), appear to fluctuate month over month with some peaks and troughs. It looks like there may be seasonal patterns or specific events causing these variations. There is an overall increasing trend in the data points over the years.



Figure 1: Online Sales Data

This graph represents monthly in-store retail sales in goods for the same time period. The values here are also in millions and exhibit more significant fluctuations compared to the online sales data. There seems to be a sharp increase or peak in sales around the start of 2020, which might indicate a seasonal effect or an event that caused an increase in in-store purchases. Both graphs provide a visual representation of sales trends over time and could be useful for analyzing the performance of online vs. in-store sales, understanding seasonal effects on retail, and planning for inventory or marketing strategies.



Figure 2: Offline Sales Data

There are two lines on the graph: one for in-store sales, which is shown in purple, and one for online sales, shown in orange. The in-store sales line starts higher and shows more variability, with peaks and troughs throughout the period. It also seems to have a slight overall decline by the end of

the period. In contrast, the online sales line starts much lower, shows growth over time, and is less volatile than the in-store sales.

This graph could be illustrating trends such as the impact of events like holidays or sales seasons on in-store versus online sales, or perhaps the general increase in online shopping over the period shown. It also suggests that while in-store sales have greater monthly fluctuations, online sales have been steadily increasing over the years.



Figure 3: Online and offline sales compare

7 Statistic analysis

7.1 Seasonal analysis

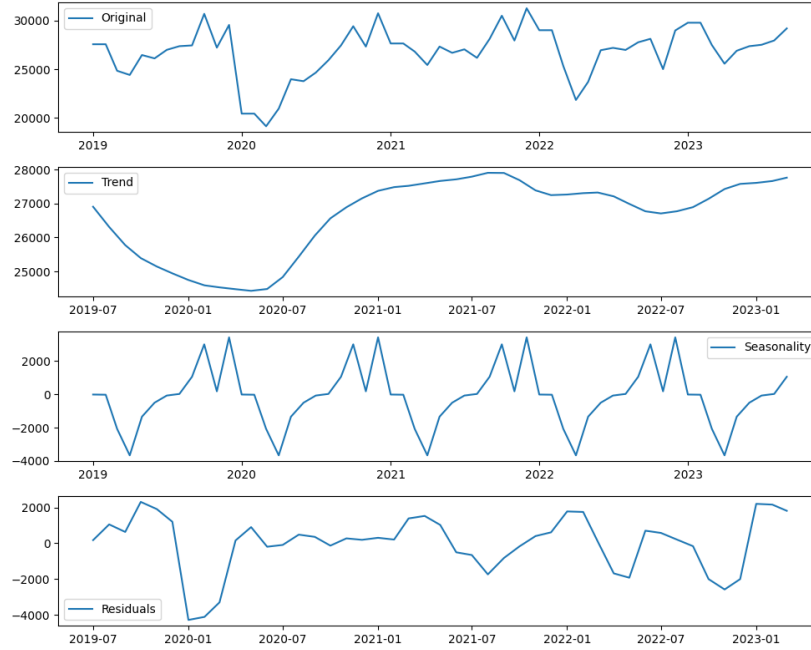


Figure 4: Offline Seasonal analysis

Offline Seasonal analysis In Figure 4, Original Data shows the original monthly sales of physical store retail. From the figure we can see that sales show certain fluctuations, which may be caused by various factors such as market demand, promotional activities or seasonal factors.

The Trend section reveals the long-term trend of the data, eliminating the effects of seasonality and other irregularities. As can be observed from the chart, the trend line showed a decline at the beginning of 2020, which was most likely related to the COVID-19 pandemic, and then the data showed a gradual recovery and continued upward trend.

Seasonality reveals seasonal patterns in data, that is, regular fluctuations that occur periodically throughout the year. The chart indicates a regular increase or decrease in sales data during specific months, indicating that sales are significantly affected by seasonality.

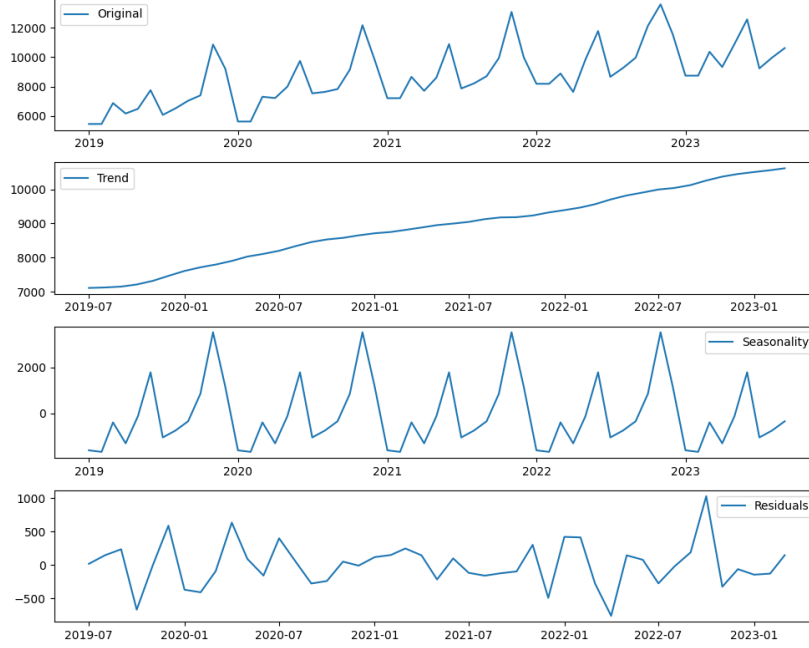


Figure 5: Online Seasonal analysis

Online Seasonal analysis Figure 5 shows the patterns observed in online sales data over several years. The original data in the top chart captures the monthly fluctuations, showing peaks and troughs that reveal how online sales volumes change over time. Notably, there is an increase in sales towards the end of each year, likely influenced by annual e-commerce promotional events in November and December. The trend component smooths out short-term irregularities and seasonal effects to elucidate the underlying long-term trend in online sales data. This trend is generally upward, indicating a consistent growth in online sales, which may be reflective of the broader expansion and adoption of e-commerce practices.

The seasonal component of the data illustrates the cyclic patterns that recur within each year. The regular occurrence of peaks and troughs suggests that online sales are significantly influenced by seasonal factors, such as holiday periods and specific sales events like Black Friday and Christmas, which traditionally see an uptick in retail activity. This seasonality is characteristic of the retail industry, where certain times of the year consistently exhibit higher or lower sales due to these well-established commercial and cultural events.

7.2 t-test

First, test the variance of the two sets of sample data: online monthly sales and offline monthly sales. At a significance level of 0.05, use the Levene statistic to determine whether the variances are equal based on the p value. If the p-value is less than the significance level, the variances of the two sets of data are considered to be unequal, and vice versa.

The results are shown as Levene Statistics: 0.6410067116098597, p-value: 0.42504157050096525 Therefore, we failed to reject the null hypothesis of equal variances and considered the variances of the two sets of data to be equal.

In the case of equal variances, We use t-test to test whether there is a statistically significant difference between the means of online and offline sales data.

Then, conduct a t test on the two sets of data, online monthly sales and offline monthly sales, to compare the means of sales. Choose a significance level of 0.05. If the test p-value is less than the significance level, the null hypothesis is rejected and there is no significant difference in the means of the two sets of sales data, and vice versa.

The results are displayed as t statistics: -41.477883180273004, p-value: 1.0659906776480866e-66. Therefore, we reject the null hypothesis and believe that there is a significant difference in the means of the two sets of data.

7.3 Regression analysis

OLS Regression Results						
Dep. Variable:	Accumulated Sales	R-squared:	0.108			
Model:	OLS	Adj. R-squared:	0.092			
Method:	Least Squares	F-statistic:	6.669			
Date:	Mon, 04 Dec 2023	Prob (F-statistic):	0.0125			
Time:	18:06:07	Log-Likelihood:	-525.45			
No. Observations:	57	AIC:	1055.			
Df Residuals:	55	BIC:	1059.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2.295e+04	1510.843	15.192	0.000	1.99e+04	2.6e+04
Accumulated Sales	0.4333	0.168	2.582	0.013	0.097	0.770
Omnibus:	5.161	Durbin-Watson:	0.967			
Prob(Omnibus):	0.076	Jarque-Bera (JB):	4.398			
Skew:	-0.669	Prob(JB):	0.111			
Kurtosis:	3.253	Cond. No.	4.14e+04			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The condition number is large, 4.14e+04. This might indicate that there are strong multicollinearity or other numerical problems.						

Figure 6: Regression results

The regression output provides various statistics about the relationship between online monthly sales (x) and in-store monthly sales (y).

The R-squared value is 0.108, meaning that approximately 10.8 of the variability in in-store sales can be explained by the variability in online sales. This is a relatively low value, suggesting that online sales are not a strong predictor of in-store sales.

The F-statistic is 6.669 with a p-value of 0.0125. This suggests that the model is statistically significant at a common alpha level of 0.05, meaning there is a statistically significant relationship between online and in-store sales.

Coefficients: The intercept is approximately 22,950 with a p-value of 0.000, indicating that it is statistically significant. This value represents the expected in-store sales when online sales are zero. The coefficient for online sales is 0.4333 with a p-value of 0.013, which is also statistically significant. This coefficient means that for every unit increase in online sales, in-store sales increase by approximately 0.4333 units.

In summary, while the model indicates a statistically significant relationship between online and in-store sales, the low R-squared value means that online sales are not a strong predictor of in-store sales. With the growth of online retail sales, the in-store retail sales will also increase slightly.

This is contrary to our guess that the growth of online sales will greatly weaken offline sales. The reason for this may be that the overall social economy has recovered after the epidemic, and both online and offline sales have increased.

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.162			
Model:	OLS	Adj. R-squared:	0.142			
Method:	Least Squares	F-statistic:	8.292			
Date:	Mon, 04 Dec 2023	Prob (F-statistic):	0.00618			
Time:	18:12:17	Log-Likelihood:	-404.17			
No. Observations:	45	AIC:	812.3			
Df Residuals:	43	BIC:	815.9			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.966e+04	2436.490	8.070	0.000	1.47e+04	2.46e+04
0	0.7825	0.272	2.880	0.006	0.235	1.331
Omnibus:		7.690	Durbin-Watson:		0.538	
Prob(Omnibus):		0.021	Jarque-Bera (JB):		7.036	
Skew:		-0.955	Prob(JB):		0.0297	
Kurtosis:		3.320	Cond. No.		7.44e+04	
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The condition number is large, 7.44e+04. This might indicate that there are strong multicollinearity or other numerical problems.						

Figure 7: Regression results

7.4 Deseasonalized regression

After removing the seasonality factor and performing the regression analysis, here's the interpretation of the results:

The R-squared value is 0.162, which indicates that the model explains 16.2 of the variance in the de-seasonalized in-store sales. This is an improvement over the previous model with seasonality, suggesting that online sales are a more significant predictor of in-store sales after removing the seasonal effect.

The F-statistic is 8.292 with a p-value of 0.00618. This suggests that the model is statistically significant, and the relationship between online sales and de-seasonalized in-store sales is not due to random chance.

The coefficient for the constant term is approximately 196,600 with a p-value close to 0. This suggests a significant intercept, which could be interpreted as the baseline level of in-store sales when online sales are zero (assuming the model is correctly specified).

The coefficient for online sales is about 0.7825 with a p-value of 0.006, indicating a statistically significant relationship between online sales and de-seasonalized in-store sales. For every unit increase in online sales, de-seasonalized in-store sales increase by approximately 0.7825 units.

8 Conclusion

In this scientific study, we leveraged data sourced from the National Bureau of Statistics of China, employing Python-based web scraping methodologies to establish a foundational dataset for our investigation. This data acquisition and processing approach enabled an extensive examination of the evolving retail trends spanning from 2019 to 2023. Through rigorous data analysis, we sought to elucidate the dynamic interplay between online and offline retail channels within the Chinese market.

The principal finding of our research is the identification of a marked shift in consumer preferences towards digital platforms, significantly impacting the traditional brick-and-mortar retail paradigm. This trend is further accentuated by the COVID-19 pandemic, which has expedited the transition towards e-commerce. Our analysis suggests that these changes in consumer behavior are not transient but indicative of a more enduring transformation, underscoring the necessity for adaptability and innovation within the retail sector.

Furthermore, our study reveals pronounced regional and urban-rural disparities in the development of e-commerce within China. These variations highlight the complex and multifaceted nature of the digital retail landscape in a diverse and vast economy. Additionally, we observed the increasing salience of seasonal trends in retail sales, which have discernible effects on both online and offline channels.

This aspect of retail dynamics underscores the importance of strategic planning and market analysis in accordance with cyclical consumer patterns.

Conclusively, our investigation posits that the ascent of e-commerce in China is fundamentally re-defining the retail landscape, which are essential for comprehending and foreseeing future developments in the retail sector. This comprehensive analysis contributes to the ongoing discourse concerning the future of retail in an increasingly digital age, emphasizing the imperative for the retail industry to adapt and innovate in response to these transformative trends.

References

- [FTZZ18] Jingting Fan, Lixin Tang, Weiming Zhu, and Ben Zou. The alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, 114:203–220, 2018.
- [LWZ19] Xubei Luo, Yue Wang, and Xiaobo Zhang. E-commerce development and household consumption growth in china. *World Bank Policy Research Working Paper*, (8810), 2019.
- [Min16] Ministry of Commerce of the People’s Republic of China. E-commerce in china, 2016. Ministry of Commerce of the People’s Republic of China, 2016. Accessed: [Insert Date of Access].